Statistical Distributions in Particle Technology

Peter J.T. Verheijen
Department of Chemical Engineering
Technical University Delft
The Netherlands

May 9, 2017

1 Basic Concepts

Properties of single particles of use in the chemical industry are eg chemical composition, size, shape, color, mechanical strength, magnetic susceptibility, electrical conductivity and mass. Each of these properties can be determined for single particles, but they will differ from particle to particle. Normally, we are not interested in the properties of single particles but in the behavior of whole sets of particles. Hence, we are not interested to know the properties of each individual particle, but the particle technologist has to describe the distribution of the particle properties.

In the jargon of a statistician, each property is a stochastic variable, X, and has a frequency distribution, $f(x)$, where

$$f(x)dx = \mathrm{P}(x < X < x + dx)\,,$$

the probability to find the variable X in the domain $[x, x+dx]$. When the property has only certain discrete values (eg color can only be red, orange, yellow, green, blue and purple; or crystal structure can only be fcc, bcc, etc.) then the frequency distribution has a set of discrete values, $f(x) = f_k$, and when the property is continuous (eg mass is a positive number; or equivalent sphere diameter is also a positive number) the function, $f(x)$, is a continuous function of x. The function $f(x)$ constitutes a valid statistical distribution if

$$f(x) >= 0 \text{ and } \int_{\infty}^{\infty} f(x)dx = 1.$$

These functions play a fundamental role in particle technology.

The details of a distribution are primarily given by its functional form. Either the distribution function, $f(x)$, is given or the cumulative distribution function, $F(x)$,

$$F(x) = \int_{\infty}^{x} f(u)du$$

It is a matter of subjective choice to use either of the two representations to gain insight in processes of relevance to particle technology.

Derived quantities describe, eg, the location or the spread of the distribution. It is for example easier to visualize an average diameter of particles then to know that the diameter is a stochastic variable with a certain distribution.

Several descriptors are given below, where $\mathrm{E}\,g(X)$ is the expected value of the function $g(X)$.

$$\mathrm{E}\,g(X) = \int_{\infty}^{\infty} g(u)f(u)du$$

Such expected value tells something about, X, the quantity that we want to describe. The most important ones are the average value, which describes the location of the variable, and a variable to describe the width of the distribution.

The following measures apply typically to mono-modal distributions. Bi-modal or multi-modal distributions are best characterized with their respective components and the relative strength that each component is present.

1.1 Location parameter

There are a few different statistics describing the position op the distribution:

mean	$x_{ave} = \mathrm{E}\, X$
median	x_{med} is value where $F(x_{med}) = 0.5$
geometric mean	$\ln(x_{geom}) = \mathrm{E}\,(\ln(X))$
mode	x_{mode} is value of x where $f(x)$ is maximum
quantile at p	x_p is value of x where $F(x_p) = p$, eg $x_{0.5} = x_{med}$.

For narrow distributions they are all nearly the same and then the mean has a slight preference because of its mathematical description. The median is very suitable for comparison with noisy data. The geometric mean is for broad distribution that look 'normal' on a logarithmic scale. The quantile is reserved for describing processes where the tail has an important role in a process such as in primary crystallization. For broad distributions in particle technology *typically* it is found that

$$x_{mode} < x_{med} < x_{geom} < x_{ave}.$$

1.2 Width parameter

The spread around a location is the second important statistic to describe a distribution:

variance	$\mathrm{Var}\, X = \mathrm{E}\,((X - \mathrm{E}\, X)^2)$
standard deviation	$\sigma = \sqrt{\mathrm{Var}\, X}$
geometric standard dev.	$\sigma_g = \sqrt{\mathrm{Var}\,(\ln(X))}$
quartile range	$Q = x_{0.75} - x_{0.25}$
range	$Q = x_{0.90} - x_{0.10}$, or equivalent.

The variance is the quantity most easily used in theoretical descriptions. The standard deviation is more easily visualized. In the case of multi-modal or strongly skewed distributions, the variance is not very suitable and then the ranges are better descriptors.

The geometric standard deviation can be substituted directly in the width parameter for the log-normal distribution, which particle technologists often use[1]. For narrow distributions this is approximately the ratio, $\sigma/x_{ave} \approx \sigma_g$.

The quartile ratio is used for comparison with robust measurements. The range is useful for quick estimations.

1.3 Shape parameter

Further a distribution is often symmetrical or is just characterized by a given asymmetry or deviation form a "normal" distribution. The following are dimensionless

[1]An annoying matter is, that particle technologist call $\sigma_{g,parttechn}$ in $\sigma_g = \ln(\sigma_{g,parttechn})$ also geometric standard deviation. This is however somewhat clumsy.

quantities, which appear often in older literature.

$$\text{skewness} \qquad \frac{\mathrm{E}\left((X-\mathrm{E}\,X)^3\right)}{\sigma^3}$$
$$\text{kurtosis} \qquad \frac{\mathrm{E}\left((X-\mathrm{E}\,X)^4\right)}{\sigma^4}-3$$

The first of these two determines the lack of symmetry around the center of the distribution. It is positive for distributions with a tail to the right. The kurtosis is a measure for peakedness. Sometimes the term "3" is not subtracted, which is mathematically more elegant. However, in this manner the peakedness is zero for a normal distribution (see below). So, in this way a measure relative to the normal distribution is obtained.

1.4 Moments

Theoretical derivations often make use of theoretical quantities, called moments, defined as

$$k^{\text{th}} \text{ moment} \qquad M_k = \mathrm{E}\left(X^k\right)$$
$$k^{\text{th}} \text{ central moment} \qquad M'_k = \mathrm{E}\left((X-\mathrm{E}\,X)^k\right)$$

By calculating M_k for $k = 0, 1, 2 \ldots$, from data or directly from the known distribution function, also a representation of the distribution function is obtained. Its parallel is the Taylor expansion to define a function. In principle, any distribution function can be expressed as a series of its moments. The higher the moment the less significant it becomes to describe the distribution function. Although this representation is rather vague at first inspection, it is a useful tool in theories predicting the change of distribution by deriving relations for the most important terms: $k = 0, 1$ and 2.

$M_0 = 1$ or, if a number distribution is used, $M_0 =$ the total number of particles. M_1 is the mean. The second moment is connected to the spread: $\mathrm{Var}\,X = M_2 - M_1^2 = M'_2$.

Skewness and kurtosis are nothing more than the higher central moments scaled by the measure for width.

2 Relevant distributions

The following distributions are normally only applied to sizes, however some have a broader applicability. A summary is for example given by Yu and Standish (1990). Below a number of distribution are described. x_0 is a scaling parameter of the stochastic variable on the order of the common location estimators. The parameters a and b are tune parameters. Where necessary, a lower and upper limit of a distribution are given by x_{min} and x_{max}. An estimation of the use of these distributions follows from an analysis of articles in the period 2000–2016[2]: log-normal distribution, 40%; Rosin-Rammler/Weibull distribution, 20%; power law, order of 30%; exponential distribution, 1%; and gamma distribution, 5%[3].

[2] This has been done with keywords search in the SCOPUS data base. A problem is that the word 'particle' could also refer to protons, electrons, etc. Secondly, the power law can be used in all kinds of contexts, not only distributions. So the numbers given are rough estimates to give an idea.

[3] In 1990-2000 these numbers were 65%, 20%, 5% and 5%

2.1 The normal and log-normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)$$

This is the result of the Central Limit Theorem in statistics and is also a good approximation in a number of cases. If $\ln(X)$ has a normal distribution, then X has a log-normal distribution which has a wide applicability. It is described in the next section.

2.2 The Rosin-Rammler or Weibull distribution

$$f(x) = a\frac{1}{x_0}\left(\frac{x}{x_0}\right)^{a-1} \exp(-(x/x_0)^a) \text{ and } F(x) = 1 - \exp(-(x/x_0)^a)$$

This was originally applied to data from crushed coal. It has a convenient form and can describe a large range of broad distributions. It is computationally convenient and it has some physical meaning to be explained in a section below. In the statistics the function is known as the Weibull distribution applied in reliability analysis.

2.2.1 Nukiyama-Tanasawa distribution

For $a = 3$ a special case of the Rosin-Rammler distribution is obtained. This has been used to describe sprays. Apparently, it was originally introduced as a phenomenological equation, but there are, apparently, also theoretical arguments for its application in the droplet fragmentation.

2.3 The power law

$$f(x) = \left(\frac{x}{x_{min}}\right)^{-a-1} \frac{a}{x_{min}} \text{ and } F(x) = 1 - \left(\frac{x}{x_{min}}\right)^{-a}$$

This is actually only valid as a local approximation for part of a certain range of the variable. It is sometimes referred to as Junge distribution. Advantage: ease of use in simple theoretical derivations.

2.4 The exponential distribution

$$f(x) = \frac{1}{x_0} \exp\left(-\frac{x - x_{min}}{x_0}\right) \text{ and } F(x) = 1 - \exp\left(-\frac{x - x_{min}}{x_0}\right).$$

This has been applied to size distributions of powders and as an approximation for large diameters of aerosols. This has a very simple form and can also be used if the details of a distribution near $x = 0$ is not very important, while at times the asymptotic approximation is only of importance to the phenomena to be described.

2.5 The Gamma function and related distributions

$$f(x) = \frac{1}{x_0\Gamma(a)}\left(\frac{x}{x_0}\right)^{a-1} \exp(-x/x_0)$$

Atmospheric aerosol data have been described by this formula and it is applied to the description of powders. For $a = 1$ the exponential distribution is obtained. The sum of k stochastic processes each described by the same exponential distribution has this distribution with $a = k$.

2.5.1 The inverse Gamma function

$$f(x) = \frac{1}{x_0\Gamma(a-1)}\left(\frac{x_0}{x}\right)^a \exp(-x_0/x)$$

Berglund and Jong (1990) used the relation to describe the distribution from growth and nucleation in a crystalliser, where the model equations included growth rate dispersion. However, there was no theoretical justification for this formula.

2.5.2 The powered Gamma function

$$f(x) = \frac{1}{x_0 b\Gamma(a)}\left(\frac{x}{x_0}\right)^{a/b-1} \exp(-(x/x_0)^{1/b})$$

gives an extra parameter to tune in order to describe a distribution. The above mentioned authors also applied this for the same problem.

2.6 The Beta distribution

$$f(x) = \frac{\Gamma(a+b)}{x_{max}\Gamma(a)\Gamma(b)}\left(\frac{x}{x_{max}}\right)^{a-1}\left(1-\frac{x}{x_{max}}\right)^{b-1}$$

has a definite maximum value and is therefore useful in defining processes with a definite upper limit eg a milling process that cuts particles to a certain size or smaller (Peleg and Normand 1986). The log-Beta distribution has been mentioned in the aerosol science. If $\ln(X/x_0)$ has a Beta distribution, then X has a log-Beta distribution.

2.7 Remarks on computability

The power law, the exponential distribution and the Rosin-Rammler distribution have all the advantage that the cumulative distribution is given in analytical form, which is easily implemented in eg a spreadsheet. The exponential distribution, the gamma distribution and the powered gamma distribution move from a single parameter (besides a location parameter) via a two parameter to a three-parameter formula. This allows for slowly introducing increasingly more refined distribution descriptions.

2.8 Special Case: Log-normal Distribution

This distribution plays an important role in particle technology. Therefore the basic properties are summarized here. The essential idea is that if

$$\frac{\ln\left(\frac{X-x_0}{x_g-x_0}\right)}{\sigma_g} \sim N(0,1),$$

a standard normally distributed quantity, then the stochastic variable X has a log-normal distribution. Often $x_0 = 0$, then the subsequent formula for the distribution, $f(x)$, of X is

$$f(x) = \frac{1}{x\sigma_g\sqrt{2\pi}} \exp\left(-\frac{\ln^2(x/x_g)}{2\sigma_g^2}\right)$$

The parameter σ_g is the geometric standard deviation, a dimensionless number. It is always quoted as such, while seldom the exponent is taken. The parameter x_g could be positive or negative, but then the stochastic variable X can take *either* positive *or* negative values.

The descriptors, given before, can here be expressed in terms of x_g and σ_g

mean	$x_{ave} = x_g \exp(\sigma_g^2/2)$
median	$x_{med} = x_g$
geometric mean	$x_{geom} = x_g$
mode	$x_{mode} = x_g \exp(-\sigma_g^2)$
quantile	x_p is expressed in terms of inverse error function
variance	$\mathrm{Var}\, X = x_g^2 \exp(\sigma_g^2)(\exp(\sigma_g^2) - 1)$
stand. deviation	$\sigma = x_g \exp(\sigma_g^2/2)\sqrt{\exp(\sigma_g^2) - 1}$
quartile range	$Q = x_{0.75} - x_{0.25}$
skewness	$(\exp(\sigma_g^2) + 2) \times \sqrt{\exp(\sigma_g^2) - 1}$
kurtosis	$\exp(4\sigma_g^2) + 2\exp(3\sigma_g^2) + 3\exp(2\sigma_g^2) - 6$
k^{th} moment	$M_k = x_g^k \exp(k^2\sigma_g^2/2)$

Vice versa, from a mean, x_{ave} and standard deviation, σ, the corresponding x_g and σ_g are determined by

$$x_g = \frac{x_{ave}}{\sqrt{1 + \sigma^2/x_{ave}^2}} \text{ and } \sigma_g = \sqrt{\ln(1 + \sigma^2/x_{ave}^2)} \simeq \frac{\sigma}{x_{ave}} \text{ for } \sigma \ll x_{ave} \qquad (1)$$

or from the first and second moment by

$$x_g = \frac{M_1}{\sqrt{M_2}} \text{ and } \sigma_g = \sqrt{\ln(M_2/M_1^2)}. \qquad (2)$$

3 Distribution presentations for particle technology

All the above considerations are of relevance to any of the possible particle properties. The most important of those is of course the size of a particle. However, there is a proliferation of methods of depicting this size. There are the following possibilities:

X = diameter, surface and volume, and

$f(x)$ = fraction of number of particles, fraction of diameters, fraction of surface and fraction of mass,

ie on the x-axis and y-axis there are respectively three and four possibilities resulting in 12 combinations. The purely mathematical description would be:

X = diameter; $f(x)$ = fraction of number of particles.

A pragmatic choice is to use

X = diameter; $f(x)$ = fraction of mass.

This is a consequence of interpreting the often used sieve measurements. The sieve mesh determines the diameter, while in fact the weight is measured. In industrial uses the output is measured in volume, so the natural choice would be to use particle volume (or weight) as the stochastic variable and $f(v)$ as volume (or weight) fractions..

Basically, there is no problem in converting from one distribution to the other

$$
\begin{aligned}
f_N(x) &= f(x) \\
f_L(x) &= \frac{x\, f(x)}{\int_0^\infty x\, f(x) dx} \\
f_S(x) &= \frac{x^2\, f(x)}{\int_0^\infty x^2\, f(x) dx} \\
f_V(x) &= \frac{x^3\, f(x)}{\int_0^\infty x^3\, f(x) dx}
\end{aligned}
$$

where the subscripts N, L, S and V mean that the distribution is expressed as respectively a number, length, surface and volume distribution with the variable, x, remaining the diameter. Eg Allen (1988) presents all relevant formulas also in the case of measurements. Changing from one stochastic variable is obtained from

$$f_N(x)dx = f_N^S(s)ds = f_N^V(v)dv,$$

where x, s and v are respectively the diameter, surface and volume. For example, the conversion from number distribution on the basis of diameters to one based on volume, $v = x^3$,[4] and vice versa can be derived from above

$$f_N^V(v) = f_N\left(\sqrt[3]{v}\right)\frac{1}{3}v^{1/3-1} \quad (3)$$

$$f_N(x) = f_N^V(x^3)3x^2 \quad (4)$$

All 12 possibilities are summarized in figure 3, which also shows that for the horizontal scale the surface and volume are presented again as a diameter.

Moments, quantiles and means are also affected by the choice of stochastic variable and the type of distribution function (number, diameter, surface or mass fraction).

Barring the ideal spheres no single particle has a well defined diameter. Maybe the particle size distribution should be a "particle average size distribution" and the average particle diameter should be the average "particle average diameter". In each case the last "average" refers to an appropriate average for one particle's irregular shape. In fact, it is the measurement method that determines the type of diameter: a Coulter Counter senses a volume so the reported diameter is an equivalent sphere diameter; a light scattering apparatus senses a projection surface so the reported diameter is an equivalent surface diameter.

4 Distributions from data

4.1 Histograms

The basic representation of a distribution is the histogram. The data in table 4.1 are those of a particle size distribution presented in three ways: the number (Figure 4.1, the height of each block is a measure for the total in the block), the number fraction (Figure 4.1, the height is a measure for the density), and the cumulative fraction (Figure 4.1).

The above is the general case with variable interval width. Quantiles follow from the cumulative fraction, eg: $x_{med} = 9$, $x_{0.25} = 5.83, x_{0.75} = 13.94$ and $Q = 8.11$. Expectation values follow from doing the correct summation. For example, the

[4] Formally, this should be multiplied by a form factor, but that is not relevant for the argument here

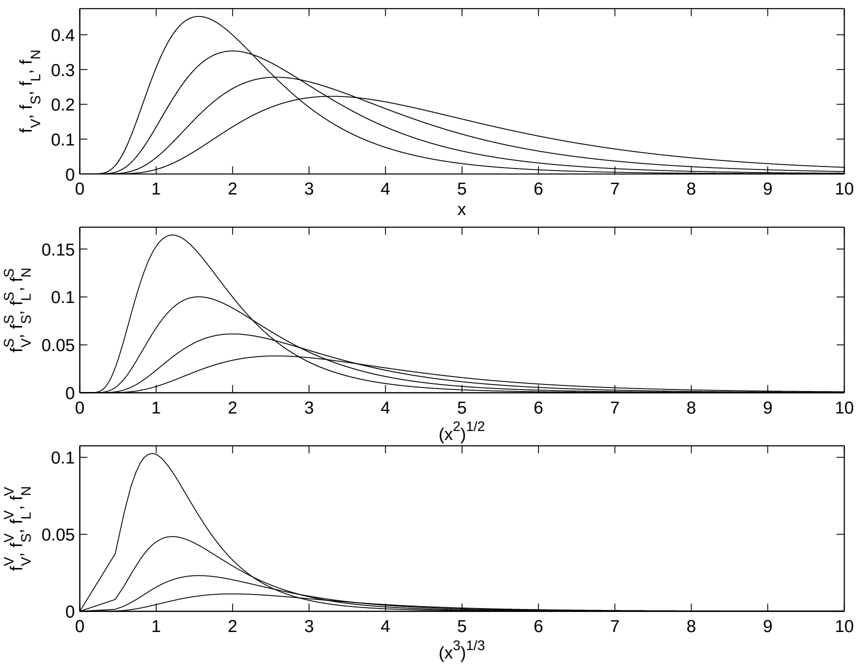

Figure 1: The twelve possibilities to present the same distribution, originally a log-normal number distribution with $x_g = 2$ and $\sigma_g = 0.5$. The maximum density is reached in order f_N, f_L, f_S and f_V for each of three cases.

calculation for the mean is performed in column 7: $x_{ave} = 10.98$, and for the geometric average in column 8: $x_{geom} = e^{2.1663} = 8.73$. Similarly, the standard deviation (7.38), geometric standard deviation (0.712), skewness (1.25) and kurtosis (1.31) can be calculated.

Here the intervals were chosen somewhat at random. For narrow or nearly normal distributions the equidistant grid is more appropriate. For broad distributions the preference is to have equidistant intervals on a logarithmic scale ie $x_{i+1} = x_i \times f$, where f is a factor > 1. A choice in this case would be $\ln(f) = \ln(10)/10 = 0.230$, because this gives 10 intervals per decade: 1, 1.26, 1.58, 2.00, 2.51, 3.16, 3.98, 5.01, 6.31, 7.94, 10, etc. In this case too the midpoint is rather the geometrical average of the upper and lower limit of an interval, $x_{midpt} = \sqrt{x_{low} x_{high}}$.

The here chosen intervals are the lower limit in case no measurement error is made, for example with a high resolution image analyzer. In the case of other measurement techniques it is the resolution of the instrument, which in fact determines the interval width.

As indicated in the introductory section, the results for the various descriptors is different, when the analysis is based on volume. The table above can be repeated with the midpoint diameter x replaced by x^3, which is a direct measure for the volume. We leave it to the reader to perform this and reproduce here only the last line of the resulting table:

Size (μm^3)	Midpt (μm^3)	Number	**Volume fraction**	Cum. fraction	Fraction per μm	Fraction ×midpt	Fraction× ln(midpt)
...							
Total		1000				23732.2	9.5772

The calculated properties of this distribution is quite different from the previous

Table 1: Data taken from Hinds (1982).

Size (μm)	Midpt (μm)	Number	**Number fraction**	Cum. fraction	Fraction per μm	Fraction $\times$midpt	Fraction$\times$ ln(midpt)
0							
4	2.00	104	0.1040	0.1040	0.0260	0.2080	0.0721
6	5.00	160	0.1600	0.2640	0.0800	0.8000	0.2575
8	7.00	161	0.1610	0.4250	0.0805	1.1270	0.3133
9	8.50	75	0.0750	0.5000	0.0750	0.6375	0.1605
10	9.50	67	0.0670	0.5670	0.0670	0.6365	0.1508
14	12.00	186	0.1860	0.7530	0.0465	2.2320	0.4622
16	15.00	61	0.0610	0.8140	0.0305	0.9150	0.1652
20	18.00	79	0.0790	0.8930	0.0198	1.4220	0.2283
35	27.50	103	0.1030	0.9960	0.0069	2.8325	0.3414
50	42.50	4	0.0040	1.0000	0.0003	0.1700	0.0150
Total		1000				10.9805	2.1663

case, which is illustrated by the cumulative distribution (Figure 4.1). Further, the equivalent diameter for the average is $\sqrt[3]{23732.2} = 28.74$ and the equivalent diameter for the geometric average is $e^{9.5772/3} = 24.35$. It is easy to the reader to extend this exercise to other parameters.

4.2 Graphs and Probability Paper

A first indication whether a certain set of data has a certain distribution is by plotting $F(x)$ in such a way that it transforms to straight line. The simple power low can be verified by plotting $f(x)$ on double logarithmic paper. Two groups of curves can be identified:

1. $F(x)$ and x lend itself to easy transformation. The power law, exponential distribution and the Rosin-Rammler distribution all can be transferred by plotting respectively $\ln(1-F(x))$ versus $\ln(x)$, $\ln(1-F(x))$ versus x (Figure 4.2 and $\ln(-\ln(1-F(x)))$ versus $\ln(x)$ (Figure 4.2). In each case the slope and/or intercept on the x-axis have a direct connection with the parameters of the respective distribution. For example, in the Rosin-Rammler distribution the slope is the parameter a and the x-axis intercept is x_0 in the formulas for the distribution.

2. The inverse of $F(x)$ can be plotted against x. This should in principle give a straight line under 45^o through the origin. However, one needs to know the distribution parameters in advance. In this way it remains a suitable albeit not useful test. In a few cases the distribution is derived from a simpler distribution, whereby a linear transformation in x or $\ln(x)$ is applied. The normal, log-normal (Figure 4.2) and exponential distributions fall in this category. The most often used is that for the normal distribution with the of inverse $F(x)$ on the y-axis and x on the x-axis.

The procedure for a normal distribution is as follows:

i Calculate at a number of points, $k = 1, 2, \ldots$, the cumulative frequency, $F(x_k)$, as a function of x_k,

ii Calculate at each point the $x_{stand,k}$ of the standard normal distribution,

iii Plot $x_{stand,k}$ against x_k.

iv Determine from the slope $1/\sigma$ and from the x-axis intercept the x_0.

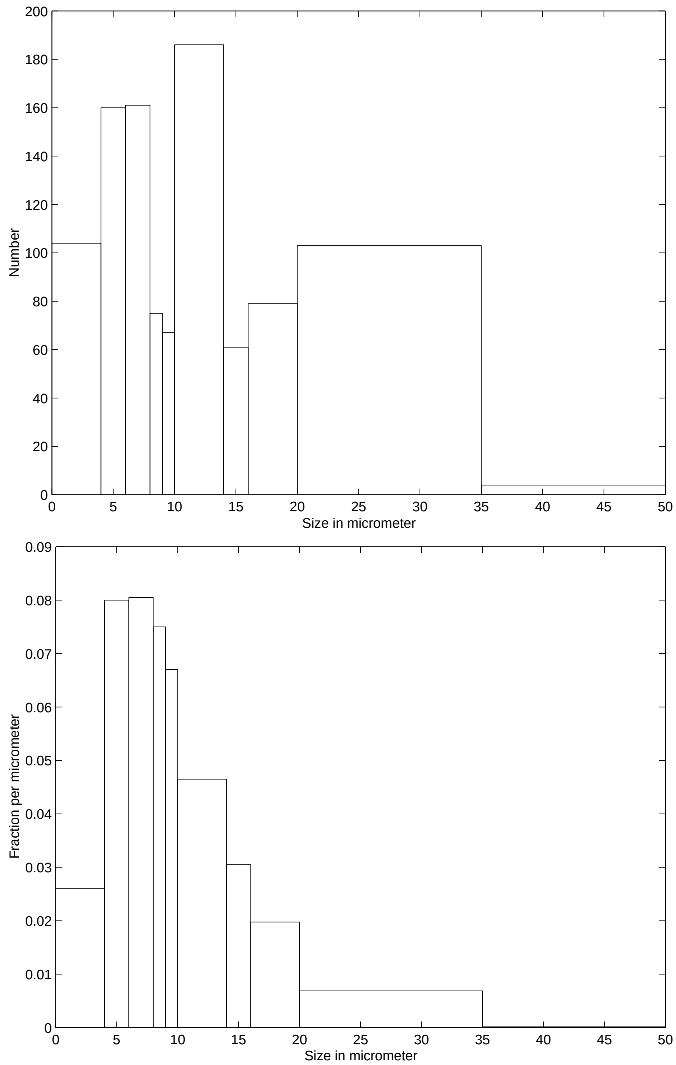

Figure 2: Histogram of (a) number and (b) density against particle size.

There exists graph paper that has the y-axis transformed in such a way that the transformation in the second step is already performed.

The usefulness of these kind of plots is limited. Most often the conclusion is that a particular set of data fits a distribution only over a certain mid-range. In any case, at the ends of a distribution the number of observations is so small, that the natural randomness of the observations should give deviations from the straight line.

4.3 Estimators

Histograms and graphs help to identify, classify and visualize the particle size distributions. Once, a choice of distribution is made, also the associated parameters should be determined with, if possible, an indication of their confidence interval. In the statistics course is shown (Dekking et al. 2004), that there are various methods for estimating the parameters in a distribution. Here, some practical approaches are presented.

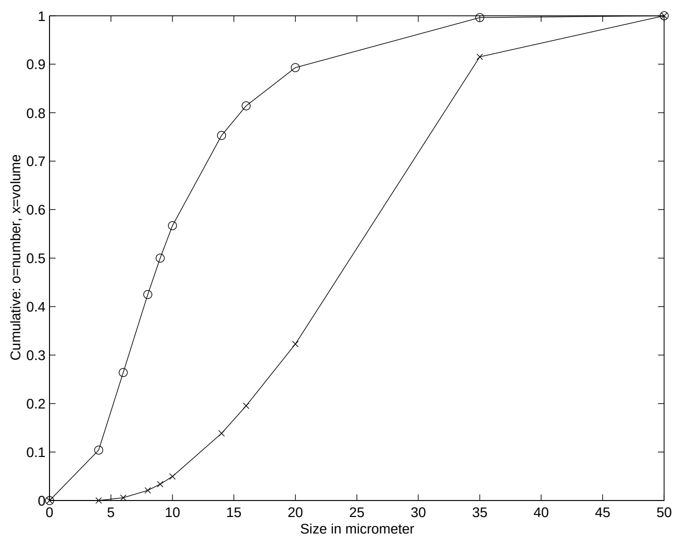

Figure 3: Cumulative distribution functions.

1. Location and width estimators have a direct relation to the distribution parameters. Equations (1) and (2) are such a relation for the log-normal distribution: x_g and σ_g are calculated from the average and standard deviation. In this case it is better to apply the geometric average and geometric standard deviation as these are direct estimators for the distribution parameters (Figure 4.3).

   A word of caution should be raised here. Sampling and measurement methods are not perfect. On the other hand it is sometimes possible that there are processes which produce for example suddenly large particles. It is therefore important to isolate the special events. These special events can then be investigated separately, while the bulk of the distribution can be described by one of the distributions described before. In those cases, it is best to apply robust estimators for location and width (median and quartile range) and relate those to the parameters of the distribution under consideration. As a rule of thumb, to isolate the 'special events' is nothing more then determining the points outside the region

   $$[x_{0.25} - 1.5Q, x_{0.75} + 1.5Q] = \frac{x_{0.25} + x_{0.75}}{2} \pm 2Q,$$

   where Q is the interquartile range, $x_{0.75} - x_{0.25}$.

2. Graphical presentation of the data leads directly to the determination of distribution parameters, but this is limited to a select number of two-parameter distributions.

3. Nonlinear regression is applied to the available data with the distribution parameters as adjustable parameters. The input data to such a program are histograms or cumulative frequency diagrams as determined. The functions to be fitted are those given above. Some pitfalls:

   (a) start values of the parameters are sometimes difficult to determine. Remedy: use the estimators of above to start the nonlinear regression.

   (b) there is tendency to fit the density function to the frequency at the midpoint of histogram intervals. Hereby, it is forgotten that the density

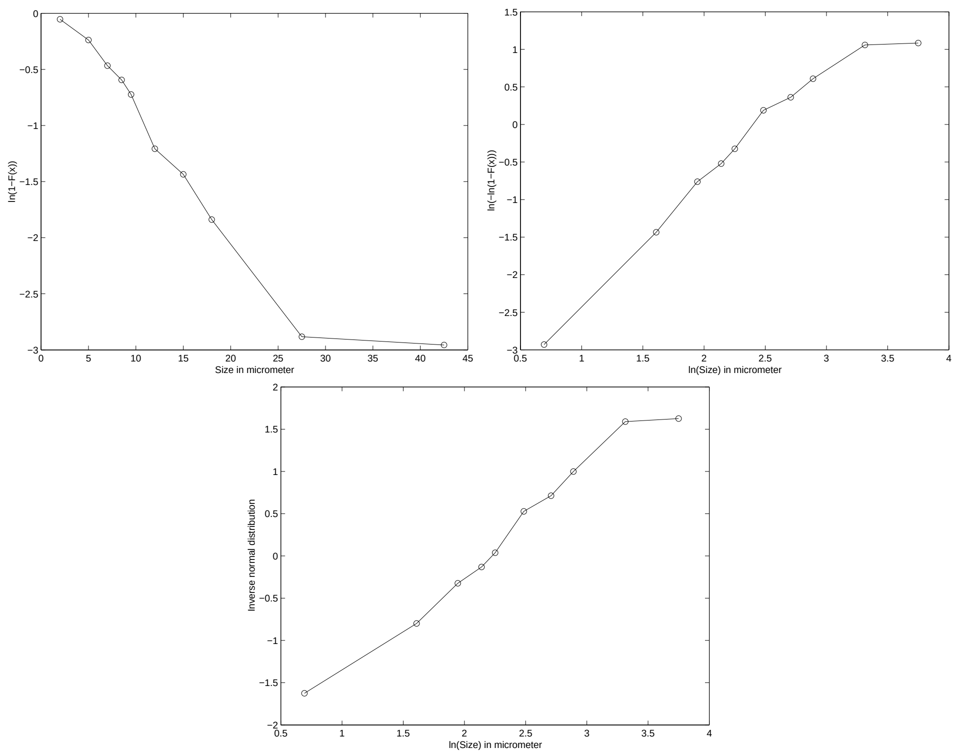

Figure 4: Histogram data tested against (a) an exponential distribution (b) a Rosin-Rammler distribution and (c) a log-normal distribution.

should be integrated over the total interval. Remedy: use the cumulative distribution and fit this with $F(x)$.

(c) the regression procedures try to fit any set of data even if they are not concordant with the expected distribution or the input data are essentially bimodal. Remedy: inspect the data beforehand to test especially on unimodality and afterwards look at the residues plot for any deviation from randomness.

(d) one of the problems here is to give a good estimate of the error in the input data. Specifically, when the number, N, of particles in each bin of a histogram is given, the expected error is derived from the Poisson process and equal to $\sqrt{N}$.

4. The maximum likelihood estimator is almost the only universal quantitative estimation method with a sound theoretical basis and suitability for the whole range of distributions used in particle technology. The idea is as follows. We are given a series observations, $x_1, x_2 \ldots x_n$ and a density function $f(x|\boldsymbol{\beta})$ where we use the vector $\boldsymbol{\beta}$ for all available parameters, eg $\boldsymbol{\beta}^t = (a, x_0)$ for a Weibull distribution. Evaluate

$$L(\boldsymbol{\beta}) = \prod_i f(x_i|\boldsymbol{\beta}) \text{ or } \ln L(\boldsymbol{\beta}) = \sum_i \ln f(x_i|\boldsymbol{\beta}),$$

where for each point the observation, x_i, is substituted and on the right hand sight we have a function in the parameters, $\boldsymbol{\beta}$, alone. When we find the parameters at the maximum of this function, we have a so-called maximum likelihood estimate,

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \ln L(\boldsymbol{\beta}).$$

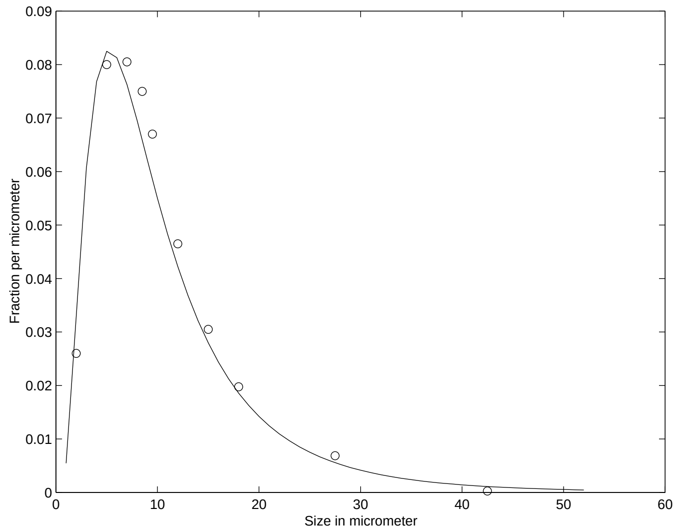

Figure 5: Histogram densities compared with the curve based on the calculated x_g and σ_g filled into the log-normal distribution.

The name comes from the fact that $L(\boldsymbol{\beta})$ is in fact a measure for the probability of the set of measurements occurring given a value for the parameters. The procedure above maximizes this probability.

4.4 Goodness-of-fit

The practice in particle technology is very simple: look at the figure and determine qualitatively whether the outcome of a certain estimation fits on the actual data. There is only two good reasons for that, namely that measurement techniques are intrinsically limited in their accuracy and that the smooth distributions given in section 2 are only approximations of a much more complex reality.

There are however, two quantitative measures:

1. The non linear regression suggested in the previous section can be evaluated on its adequacy. Once the final fit has been reached the standard methods of establishing goodness-of-fit eg with a χ^2-test on the sum of squared residuals can be performed to test whether the adequacy of a distribution model.

2. Look at the distance between the experimental cumulative distribution curve and the estimated or theoretical curve. This deviation has a known distribution depending on the number of particles observed; and use the knowledge to reject that the measured distribution is well described given a prior agreed significance level (e.g. if the probability of a large deviation is found to be smaller than 5 % then we will not reject the hypothesis). This is called the Kolmogorov-Smirnov test, which can also be adapted to test whether two measured distributions significantly differ or not. A variant with larger resolving power is the Anderson-Darling test, which weights the deviation – by $\sqrt{F(x)(1-F(x)}$ – depending on the position in the distribution. Recipes for these two tests are given by Press et al. (2007).

5 Distributions from theory

The science of particle technology is there essentially to predict the distributions that govern the properties of particles as derived from the basic particle production and destruction processes. This is still an open field of investigation. Hereunder a few cases are shown, where a distribution is derived from basic principles.

5.1 Log-normal distribution

The central limit theorem in statistics states that if $X_k, k = 1, 2, \ldots n$, are independent random variables the variable equal to the sum of X_k will approach a normal distribution for large n. Similarly, if $\ln(X_k), k = 1, 2, \ldots n$, are independent random variables the variable equal to the sum of $\ln(X_k)$ approaches a normal distribution for large n. The last statement is equivalent to stating that, if $X_k, k = 1, 2, \ldots n$, are independent positive random variables the stochastic variable equal to the product of X_k will approach a log-normal distribution for large n.

In conclusion, if a particle property, such as size, is the result of a large number of multiplicative processes than this property will tend to a log-normal distribution. Particle size, observed under various processes, is an example where this idea applies. So a number of experimentally observed particle size distributions have the properties of a log-normal distribution: it cannot predict negative diameters (!), and it is strongly skewed to the right.

5.2 Weibull or Rosin-Rammler distribution

Break-up of particles, epidemics and reliability of complicated processes have some things in common. These are all processes that can take place along different lines.

A process breakdown takes place often because of a cascade of events. Somewhere a fault takes place, eg malfunctioning pump or a sticking valve. The control system tries to correct it, but then other parts of the system gets over loaded. Finally, the system suffers a stop because the weakest link will break. An epidemic has similar features. It will start and pass on the infection along different lines. Different people are more or less susceptible and a branching process takes place spreading the virus to more people, or the infection ends as no victims are susceptible. It has been found empirically

In particles, there are breakage processes. These can be considered as branching processes as well. An interesting approach originates from Brown and Wohletz (1995). They noted that the fractures could be related to fractal processes, where they introduced a fractal dimension, D_f, which has an upper limit of 3 for three-dimensional objects. The postulate was then that fracture could take place from bigger particles of size, x', to smaller particles of size x with a probability function

$$b(x' \to x) = b(x', x) \propto x^{-D_f} \text{ with } x' > x.$$

Subsequently, they noted that the real distribution is the result of a stationary state where

$$f(x) = \int f(x')b(x', x)dx',$$

which has the solution $f(x) = \exp(-(x/x_0)^a)$ with $a = D_f$, the Weibull or Rosin-Rammler distribution (section 2.2) with as upper limit for of 3 just that of the Nukiyama-Tanasawa distribution in section 2.2.1.

The original paper of Rosin and Rammler of 1933 used the weakest link approach and postulated that there would be risk factor proportional to $(S/S_0)^a$, where S is a strength. So, S_0, therefore x_0, and a are material properties.

5.3 Power law

In the description of entities connected by weak or strong links it is often found that the number of links or other similar quantities follow power laws. This is particularly true in the area of self-similarity where structure are repeated on various scales. This also applies to particles that could be considered of consisting of different elements linked to each other. The power law here has then a root in the self-similarity of shapes.

5.4 Clusters of Molecules near Saturation

Consider a vapor of molecules near its saturation pressure, p_s. In this case there will be an equilibrium between the non-condensed molecules and the clusters. There is a rate of evaporation and absorption when single molecules hit a cluster. A straightforward calculation (Friedlander 1977) that the two processes balance at a ratio between the number of clusters with i and $i-1$ molecules of

$$\frac{n_i}{n_{i-1}} = \frac{p_1}{p_s} \exp\left(-\frac{\pi \sigma_S x_1^2}{3kT\sqrt[3]{i}}\right),$$

where σ_S is the surface tension, p_1 the vapor pressure and x_1 is the size of a single monomer. The number distribution then follows from

$$\frac{n_i}{n_1} = \frac{n_i}{n_{i-1}} \frac{n_{i-1}}{n_{i-2}} \cdots \frac{n_2}{n_1} = \left(\frac{p_1}{p_s}\right)^{i-1} \exp\left(-\frac{\pi \sigma_S x_1^2}{3kT} \sum_{j=2}^{i} \frac{1}{\sqrt[3]{j}}\right),$$

The summation over j can be approximated by an integral, so that

$$n_i = \frac{p_s}{kT} \left(\frac{p_1}{p_s}\right)^i \exp\left(-\frac{\pi \sigma_S x_1^2}{2kT} i^{2/3}\right),$$

As long as i increases and $p_1 < p_s$, the number distribution decreases roughly as an exponential distribution in $S \propto i^{2/3}$ or the surface area of each cluster. If we were to translate this into a distribution as defined earlier, the natural stochastic variable to express this in is, $v = x_1^3 i$. So,

$$f_N^V(v) \propto \exp\left(-\frac{\pi \sigma_S}{2kT} v^{2/3}\right)$$

and with equation (4),

$$f_N(x) \propto x^2 \exp\left(-\frac{\pi \sigma_S}{2kT} x^2\right)$$

In fact, this equation can be related to the Boltzmann distribution, where the energy term is replaced by the energy due to the presence of a surface term. When $p_1 > p_s$ then the interesting situation occurs that the number distribution will increase above a certain critical diameter,

$$x_{crit} = \frac{4\pi \sigma_S x_1^3}{6kT \ln(p_1/p_s)},$$

which is the point with a minimum in the size distribution. Above this point the clusters continue to grow.

6 Conjecture

In the above a number of processes are named and connected to specific distribution. We can generalize this without giving a proof. [Literature needed!]. For solid particles:

- When solid particles are mostly the result of breakage processes, then apply the Rosin-Rammler/Weibull distribution.
- When solid particles are mostly the result of agglomeration or growth processes they tend to follow a power law over a considerable range.
- When both previous situations are valid the log-normal distribution becomes a good approximation.

For liquid particles this does not apply.

7 Distributions applied

7.1 Electrostatic spray painting

It is easily shown that a cloud of electrically charged droplets can be described by a simple equation

$$m\mathbf{a} = q\mathbf{E} + \mathbf{F}_D + m\mathbf{g},$$

where the used variables have their usual meaning. It is nothing more then a the effect of a force consisting of an electrical, friction and a gravitation force. Suppose that a cloud of particles is substantially stationary then the equation above can be integrated over all particle sizes. We can then note the following relations:

- the mass, m, will be proportional to x^3, and therefore proportional to M_3,
- the electric charge, q, is proportional to x^2 and with that M_2,
- the friction force, F_D, is proportional to M_1 in the laminar region, and
- the gravitational force is proportional to the mass and therefore M_3.

The equation can then be rewritten to

$$M_3\mathbf{a} = C_E M_2 \mathbf{E} + C_D M_1 \mathbf{F}'_D + M_3\mathbf{g},$$

Normally, modelling studies in the literature do not take into account the fact that the moments will change with the width of the distribution. If, for example, we assume a log-normal distribution with geometric standard deviation σ_g. Then the equations in section (2.8) can be applied here for the moments. The result is that the acceleration

$$\mathbf{a} = C_E \exp(-2.5\sigma_g^2)\mathbf{E} + C_D \exp(-4\sigma_g^2)\mathbf{F}'_D + \mathbf{g},$$

will depend on the width of the distribution, such that the friction force and the electric force will become decreasingly less important if related to the gravitational force.

The analysis given above assumes, that the distribution will not change within a cloud, which of course is not true in general. Of course, the actually phenomena are considerably more complicated, but this first approach gives some physical insight in the tendencies to expect.

7.2 Aerosol dynamics

The primary distribution function used in this field is $n(v,t)$: $n(v,t)dv$ = the number of particles at time t in volume fraction $(v, v+dv)$. The general equation describing nucleation, condensation and coagulation is

$$\frac{\partial n}{\partial t} + \frac{\partial Gn}{\partial v} - I(v^*)\delta(v - v^*) =$$
$$\frac{1}{2}\int_0^v \beta(v-u,u)n(v-u,t)n(u,t)du - n(v,t)\int_0^\infty \beta(v,u)n(u,t)du$$

where G = condensation term, I = a nucleation term (implying that particles with v^* are created) and β = a collision frequency. These three functions need to be defined beforehand and will depend on the details of the particle-particle interactions. The equation is a partial differential equation which can only be solved numerically.

An approximate answer is obtained by focusing on the moments. The k-th moment is introduced in the formulas by multiplying the equation above left and right with x_k and integrating over all v. If only a few moments are significant in the problem, only a few equations need to be retained and they form a set of ordinary differential equations (ODE). These ODE's are nowadays solved with ease.

However, in order to reconstruct the distribution function a suggestion must be made about the actual functional form of the particle size distribution. An appropriate form is the log-normal distribution for the diameter. The moments derived from solving the ODE's are then translated to the basic parameters, σ_g and x_g or v_g, describing the diameter distribution as a function of time.[5]

[5] There is one thing to watch here. The equations require $f_N^V(x)$ in stead of $f_N(x)$. It is easy to show that the distribution based on the volume is also a log-normal distribution with geometric mean $v_g = \pi/6x_g^3$ and geometric standard deviation equal to three times the geometric standard deviation for the distribution based on the diameter.

A Statistics with the spreadsheet Excel

Spreadsheets are very useful tools to deal with data. In this section, some of those facilities of the Excel spreadsheet are described. A word of caution is required. These tools are very error prone because it depends on the user to organize the calculations via references to cells. Once a spreadsheet becomes slightly complex, an error in these references is not easily spotted[6]. Secondly, Excel is a commercial program, which means that the user is always provided with an answer even when not accurate enough or plainly wrong. This last will not occur with the simple calculations such as an average, but the more complex optimization does not always lead to the real optimum. A certain scepticism and result testing is imperative.

Excel knows add-ins. These are units that extend the functionality of this spreadsheet. The standard *installation* of Excel does not include some important ones. The standard *distribution*, however, contains the *Analysis ToolPak* for statistical analysis and the *Solver Add-in* for optimization problems. Both should be installed, which can be done through the TOOLS/ADD-IN menu option.

It is assumed here, that at least Excel of Office 2013 is used, and that later versions retain the same functionality.

A.1 Descriptive statistics

The functions in table 2, to be substituted in a cell, can be directly applied to a set of cells. A number of the statistics are recorded directly through the Analysis Tool-

Table 2: Excel functions for descriptive statistics

Location statistics	
AVEDEV	Average of absolute deviations from mean (not equal to the median absolute deviation, MAD).
AVERAGE	Mean, $\bar{x}$.
GEOMEAN	Geometric mean, $\exp\left(\overline{\ln x}\right)$, which is sometimes useful
HARMEAN	Harmonic mean, $1/(\overline{1/x})$
MEDIAN	Median, or PERCENTILE(..,50)
PERCENTILE	k-th percentile of values in a range
TRIMMEAN	Average excluding fraction extreme data
Variability statistics	
DEVSQ	Sum of squares of deviations from mean
KURT	Estimator for kurtosis, $\mathrm{E}\,(X-\mathrm{E}\,X)^4-3$
QUARTILE.INC	Estimator for quartile
SKEW	Estimator for skewness, $\mathrm{E}\,(X-\mathrm{E}\,X)^3$
STDEV.S	Standard deviation
STDEV.P	Standard deviation of population, STDEV.P=STDEV.S$\times\sqrt{(n-1)/n}$
VAR.S	Estimator for variance, $\mathrm{E}\,(X-\mathrm{E}\,X)^2$
VAR.P	Calculates variance based of population, VAR.P=VAR.S$\times(n-1)/n$
Miscellaneous statistics	
CORREL	Correlation coefficient between two data sets
COVARIANCE.P	Covariance, the average of products of paired deviations
FREQUENCY	Frequency distribution as a vertical array – histogram in table form.

[6]There are reports that 30% of the spreadsheets contain at least one error.

Pak add-in, which is reached via the menu option TOOLS/DATA ANALYSIS/DESCRIPTIVE STATISTICS. This gives a momentary output on the basis of the then present array of data. When the data in the array changes, the menu action has to be repeated to update the result. This is unlike the functions above, that are instantaneously – or after F9 – recalculated.

The correlation and covariances for more than two sets of data are calculated with the Analysis ToolPak add-in, the menu option TOOLS/DATA ANALYSIS/CORRELATION and ../COVARIANCE. In this manner, Excel calculates complete correlation and covariance matrices.

A histogram with some extra options are also available through the Analysis ToolPak add-in, namely the menu option TOOLS/DATA ANALYSIS/HISTOGRAM. However, the graphical representation does not conform to the presentation of the histogram as an estimator for a density as is done in the world of statisticians. Excel pleases the managers by making the charts more representative for business presentations.

A.2 Statistical distributions

In the standard Excel implementation, a number of distributions can be called with simple functions, which are listed in table 3. In case the word 'DIST' appears in the function name, it implies that the density function or sometimes the cumulative distribution is calculated for a realization of the stochastic variable. In case the word 'INV' appears, Excel calculates the realization of the variable given a cumulative probability.

Table 3: Excel functions for statistical distributions. Those in **BOLD** are of special interest to particle technologists.

BETA.DIST, BETA.INV	Beta probability function
BINOM.DIST, BINOM.INV	Binomial distribution probability, smallest value for which cumulative distribution $\leq$ a criterion
CHISQ.DIST, CHISQ.INV	χ^2-distribution
EXPON.DIST	Exponential distribution
F.DIST, F.INV	F-distribution
GAMMA.DIST, GAMMA.INV	Gamma distribution
HYPGEOM.DIST	Hypergeometric distribution
LOGNORM.DIST, LOGNORM.INV	Log-normal distribution $\ln(X) \sim N(\mu, \sigma^2)$
NORM.DIST, NORM.INV	Normal cumulative distribution
POISSON.DIST	Poisson distribution
T.DIST,T.INV	Student's t-distribution
WEIBULL.DIST	Weibull distribution

The functions with 'INV' calculate the critical values and the confidence intervals. Note that also a CONF function is defined in Excel, but that is limited to a normal distribution.

The other functions calculate the densities, $f(x)$, or the cumulative distributions, $F(x)$. The P-value$= 1 - F(x)$.

A.3 Linear regression

The fitting of straight line can be obtained by the functions INTERCEPT and SLOPE or by adding the trend line in graphs. The straight line belongs to the general class of linear models to where the LINEST function can be applied. See the Excel help function on how to use it. It gives the coefficients with standard errors and the estimated measurement error. Also numbers are given for hypothesis testing to establish whether the linear model is significant.

LINEST is a very useful function which is better to use than the often used copying of trend line equation in the graph.

A.4 Optimization

Many – including statistical – problems are formulated as optimization problems. When the Solver Add-in is installed, Excel can perform general optimizations as well, via the menu option TOOLS/SOLVER[7] When choosing this, a dialog window comes up with three important inputs to give

- the option to minimize, maximize or go for a certain value of the objective,
- one or more cells to be changed by the solver to reach the objective, and
- a single cell reference containing the objective.

If need be, some extra equality or inequality constraints can be added, which are applied to cell references.

As an example, we take nonlinear regression, which is a minimization problem with the parameter estimates as adjustable variables in order to find the optimum sum of squares,

$$SS(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \Big((y_i - \eta(x_i, \hat{\boldsymbol{\beta}}))/\sigma_i \Big)^2 . \tag{5}$$

Somewhere in the spreadsheet we reserve some cells for $\hat{\boldsymbol{\beta}}$. We enter a vector with the set points $\mathbf{x}$ and calculate $\boldsymbol{\eta}(\mathbf{x}, \hat{\boldsymbol{\beta}})$ in a separate vector of cells. Next we enter the observed data $\mathbf{y}$ and calculate the differences $\mathbf{y} - \boldsymbol{\eta}(\mathbf{x}, \hat{\boldsymbol{\beta}})$ with the Excel function SUMSQ in a certain cell. This last cell, the objective, and the cells for $\hat{\boldsymbol{\beta}}$, the adjustable cells, are passed to the solver dialogue. It remains to press the button SOLVE here, *if* the start values of $\hat{\boldsymbol{\beta}}$ were well chosen.

Optimization is an iterative process, which can be controlled by some options under the OPTIONS button in the solver dialogue. The default options of Excel are aimed at getting a result. Better options are to set 'Convergence at 10^{-8}, 'Automatic Scaling' on 'on', estimates (referring to an intermediate optimization step) on 'quadratic' and the derivatives on 'central'. The default search algorithm is 'Newton'.

Fortunately, most statistical problems are well behaved and have slowly varying objectives.

Optimization problems in statistics are given in table 4. The advantage is that the problem formulation is much more flexible. Additional requirements can easily be incorporated. For example, a non-negativity constraint on parameters in engineering models is quite common, and an extra dependence between some parameters can easily taken along in the constraint equations. In general the combination of a spreadsheet with a solver is very versatile.

[7] Solving equations and optimizing an objective have mathematically and numerically much in common. Software, such as Excel but also MatLab, makes full use of this, and can therefore be used for both purposes. Therefore the name SOLVER.

Table 4: Optimization in statistics

	Type	Objective	Adjustable variables
Maximum Likelihood Estimator	max	Likelihood function	Parameters of a distribution
Weighted nonlinear regression	min	$\sum\left((y_i - \eta_i(\hat{\boldsymbol{\beta}}))/\sigma_i\right)^2$	Model parameters
D-optimality	min	$\det\left(\mathrm{Var}\,\hat{\boldsymbol{\beta}}\right)$	Set points $\mathbf{x}$

The disadvantage is that a final statistical analysis, especially in the case of nonlinear regression, lacks. Of course, doing the analysis 'by hand' can circumvent this. There are add-ins developed that do nonlinear regression including the statistical analysis.

References

T. Allen. *Particle Size Measurement.* Chapman, 1988.

K.A. Berglund and E.J. De Jong. The calculation of growth and nucleation kinetics from msmpr crystallizer data including rate dispersion. *Separations Technology*, 1:38–45, 1990.

Wilbur K. Brown and Kenneth H. Wohletz. Derivation of the weibull distribution based on physical principles and its connection to the rosin-rammler and lognormal distributions. *Journal of Applied Physics*, 78:2758–2763, 1995.

F.M. Dekking, C. Kraaikamp, H.P. Lopuhaä, and Ludolf E. Meester. *Probability and statistics for the 21st century.* Technical University Delft, 2004.

S.K. Friedlander. *Smoke, dust, and haze.* Wiley, 1977.

W.C. Hinds. *Aerosol technology.* Wiley, 1982.

Ludolf E. Meester and Peter J.T. Verheijen. *Statistiek van het proefopzetten.* Technical University Delft, 2004.

M. Peleg and M.D. Normand. Simulation of size reduction and enlargement processes by a modified version of the beta distribution. *AIChe J*, 32:1928–1930, 1986.

William A. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical recipes: The art of scientific computing.* Cambridge University Press, 3rd edition, 2007.

J. Seville, U. Tžün, and R. Clift. *Processing of Particulate Solids.* Chapman, 1996.

A.B. Yu and N. Standish. A study of particle size distributions. *Powder Technology*, 62:101–118, 1990.

B Homework task

B.1 Small questions

1. Repeat - as far as possible - the table on page 6 for
   (a) The exponential distribution.
   (b) The Weibull distribution.

B.2 Seville et al. (1996)

1. The following data are, almost completely, taken from Seville et al. (1996)

Lower Bound	Upper Bound	Count Frequency
1	5	39
5	10	175
10	20	348
20	30	187
30	40	112
40	60	89
60	80	27
80	100	13
100	150	8
150	200	2

   Assume all measurements in nanometers.

   (a) Characterize the distribution presented by the data above.
   (b) Compare these data with various distributions [there are several ways suggested during the lectures] and choose - with arguments - the best fitting distribution.
   (c) Check whether your fitted distribution changes if for the midpoint calculation not the arithmetic mean of upper and lower bound but the geometric mean is chosen.
   (d) Calculate $M_6 = \mathrm{E}(x^6)$, which is a useful quantity relevant for light diffraction measurements with small particles.

   Do all calculations with Excel, and submit your spreadsheet together with the answers.

2. Given is that in [0,1] μm is 0.50 of the number fraction of the particles, and between [1,2] μm is the other half. What is the volume fraction in the two intervals?[8]

3. There is a log-normal distribution with $x_g = 0.42\mu m$ and $\sigma_g = 2.0$. Calculate: x_{ave}, σ, M_3 (report $\sqrt[3]{M_3}$), and M_6 (report $\sqrt[6]{M_6}$).

4. One apparatus found that $\mathrm{E}(X^6) = 65.8 \times 10^{-6}$ and another apparatus found $\mathrm{E}(X^3) = 314 \times 10^{-6}$. Find the x_g and σ_g of the associated log-normal distribution

[8] Answer: respectively 0.036 and 0.964. A rather big difference!

Contents

1 Basic Concepts **1**
- 1.1 Location parameter . . . 2
- 1.2 Width parameter . . . 2
- 1.3 Shape parameter . . . 2
- 1.4 Moments . . . 3

2 Relevant distributions **3**
- 2.1 The normal and log-normal distribution . . . 4
- 2.2 The Rosin-Rammler or Weibull distribution . . . 4
  - 2.2.1 Nukiyama-Tanasawa distribution . . . 4
- 2.3 The power law . . . 4
- 2.4 The exponential distribution . . . 4
- 2.5 The Gamma function and related distributions . . . 4
  - 2.5.1 The inverse Gamma function . . . 5
  - 2.5.2 The powered Gamma function . . . 5
- 2.6 The Beta distribution . . . 5
- 2.7 Remarks on computability . . . 5
- 2.8 Special Case: Log-normal Distribution . . . 5

3 Distribution presentations for particle technology **6**

4 Distributions from data **7**
- 4.1 Histograms . . . 7
- 4.2 Graphs and Probability Paper . . . 9
- 4.3 Estimators . . . 10
- 4.4 Goodness-of-fit . . . 13

5 Distributions from theory **14**
- 5.1 Log-normal distribution . . . 14
- 5.2 Weibull or Rosin-Rammler distribution . . . 14
- 5.3 Power law . . . 15
- 5.4 Clusters of Molecules near Saturation . . . 15

6 Conjecture **15**

7 Distributions applied **16**
- 7.1 Electrostatic spray painting . . . 16
- 7.2 Aerosol dynamics . . . 17

A Statistics with the spreadsheet Excel **18**
- A.1 Descriptive statistics . . . 18
- A.2 Statistical distributions . . . 19
- A.3 Linear regression . . . 20
- A.4 Optimization . . . 20

B Homework task **22**
- B.1 Small questions . . . 22
- B.2 Seville et al. (1996) . . . 22